On averaging principle for diffusion processes with null-recurrent fast component

R. Khasminskii\textsuperscript{a,*}, N. Krylov\textsuperscript{b,1}

\textsuperscript{a}Department of Mathematics, Wayne State University, Detroit, MI 48202, USA
\textsuperscript{b}School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA

Received 10 April 2000; received in revised form 27 October 2000; accepted 27 October 2000

Abstract

An averaging principle is proved for diffusion processes of type $(X_\varepsilon(t), Y_\varepsilon(t))$ with null-recurrent fast component $X_\varepsilon(t)$. In contrast with positive recurrent setting, the slow component $Y_\varepsilon(t)$ alone cannot be approximated by diffusion processes. However, one can approximate the pair $(X_\varepsilon(t), Y_\varepsilon(t))$ by a Markov diffusion with coefficients averaged in some sense. © 2001 Elsevier Science B.V. All rights reserved.

MSC: primary 60J60; secondary 60B10

Keywords: Averaging principle; Null-recurrent diffusion; Arcsine law; Homogenization

1. Introduction

The problem of limit behavior of a slow component of multidimensional Markov diffusion process has many applications in physics, biology and other areas. One possible setting of this problem is as follows. Assume that we are given a Markov diffusion process $(X_\varepsilon(t), Y_\varepsilon(t))$ consisting of two components $X_\varepsilon(t)$ and $Y_\varepsilon(t)$ and depending on the parameter $\varepsilon$ which tends to zero. Then we are interested in what happens if, as $\varepsilon \to 0$, $X_\varepsilon(t)$ changes faster and faster in time and $Y_\varepsilon(t)$ lives in the same time scale for all $\varepsilon$.

This problem was originally considered in Khasminskii (1968). Roughly speaking, the main result of Khasminskii (1968) can be described as follows. Let the generator $L_\varepsilon(x, y)$ of the process $(X_\varepsilon(t), Y_\varepsilon(t))$ have the form

$$L_\varepsilon(x, y) = \varepsilon^{-2}L_1(x, y) + \varepsilon^{-1}L_2(x, y) + L_3(x, y),$$

\textsuperscript{*}Corresponding author.

\textsuperscript{1}Partially supported by NSF Grant DMS 9971608.

\textsuperscript{2}Partially supported by NSF Grant DMS-9876586.

0304-4149/01 - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0304-4149(00)00097-1
where
\[
L_1(x, y) = \sum_{i,j=1}^{l_1} a_{ij}^{(1)}(x, y) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{l_1} b_i^{(1)}(x, y) \frac{\partial}{\partial x_i},
\]
\[
L_2(x, y) = \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} a_{ij}^{(2)}(x, y) \frac{\partial^2}{\partial x_i \partial y_j},
\]
\[
L_3(x, y) = \sum_{i,j=1}^{l_2} a_{ij}^{(3)}(x, y) \frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^{l_2} b_i^{(3)}(x, y) \frac{\partial}{\partial y_i}.
\]

Suppose that, for any \( y \), the Markov process \( X^{x,y}(t) \) with generator \( L_1(x, y) \) and satisfying \( X^{x,y}(0) = x \) is ergodic. Assume the density of its stationary distribution exists and denote it by \( \rho(x, y) \). Also, denote
\[
\tilde{L}_3(y) = \int L_3(x, y) \rho(x, y) \, dx
\]
and let \( \tilde{Y}^y(t) \) be the Markov process starting at \( y \) with generator \( \tilde{L}_3 \). Under some additional conditions on the coefficients, guaranteeing, in particular, compactness of the family of measures generated by \( Y_{\varepsilon}^{x,y}(t) \) for \( \varepsilon \to 0 \) and weak uniqueness of the process \( \tilde{Y}^y(t) \), the averaging principle for the slow component was proved in the following form: \( Y_{\varepsilon}^{x,y}(t) \xrightarrow{\text{distr}} \tilde{Y}^y(t) \) for any \( x, y \).

More precisely, the averaging principle was proved in Khasminskii (1968) (see also Papanicolaou et al., 1977) under a slightly less restrictive assumption than ergodicity of \( X^{x,y}(t) \). Namely, it is enough to assume existence for all \( x, y \), non-randomness, and independence of \( x \) of the (a.s.) limit
\[
\lim_{T \to \infty} \frac{1}{T} \int_0^T L_3(X^{x,y}(t, \omega), y) \, dt = \tilde{L}_3(y).
\] (1.1)

Later the result of Khasminskii (1968) was extended in several papers and monographs, see e.g. Papanicolaou et al. (1977), Gikhman and Skorokhod (1982), Skorokhod (1989) and Freidlin and Wentzell (1998).

The subject of the present paper is more delicate: we study the limit behavior of the slow component in the situation in which \( X^{x,y}(t) \) is a null-recurrent one-dimensional diffusion process for each \( y \), and the limit in (1.1) does not exist. We prove that under appropriate conditions, the process \( (\varepsilon X_{\varepsilon}(t), Y_{\varepsilon}(t)) \) converges weakly to a limit \( (\tilde{X}(t), \tilde{Y}(t)) \), as \( \varepsilon \to 0 \). In comparison with the ergodic case, the essential difference is the fact that \( \tilde{Y}(t) \) is not Markovian, and only the pair \( (\tilde{X}(t), \tilde{Y}(t)) \) is a Markov process with discontinuous at \( x = 0 \) drift and diffusion coefficients. These coefficients are evaluated using some sort of averaging the coefficients in the original system.

A caricature of the situation we are concerned with can be given by \( Y_{\varepsilon}(t) = \int_0^t \sigma w_0(s)/\varepsilon \, dw_1(s) \), where \( w_0(t), w_1(t) \) are independent one-dimensional Wiener processes. It is easy to understand that \( Y_{\varepsilon}(t) \) converges in distribution as \( \varepsilon \) to the process
\[
\int_0^t (\sigma^+ I_{w_0(s) > 0} + \sigma^- I_{w_0(s) < 0}) \, dw_1(s)
\]
if the limits \( \sigma^\pm := \lim_{x \to \pm \infty} \sigma(x) \) exist. But what happens if at least one of them does not exist or the process \( Y_s(t) \) is not so simple? This is the question we are trying to answer.

2. Assumptions and the main result

We consider a Markov process \( Z_s(t) = (X_s(t), Y_s(t)) \) with a fast component \( X_s(t) \in \mathbb{R}^1 \) and a slow component \( Y_s(t) \in \mathbb{R}^d \). Assume that this process is a solution of Itô’s differential equations

\[
\begin{align*}
\mathrm{d}X_s(t) &= \varepsilon^{-1} \varphi(X_s(t), Y_s(t)) \mathrm{d}w(t), \\
\mathrm{d}Y_s(t) &= b(X_s(t), Y_s(t)) \mathrm{d}t + \sigma(X_s(t), Y_s(t)) \mathrm{d}w(t), \\
X_s(0) &= x, \quad Y_s(0) = y.
\end{align*}
\]

(2.1)

Here \( \sigma(\ldots) = (\sigma_{ij}(\ldots)) \) is \( d \times k \) matrix,

\[
\varphi(\ldots) = (\varphi_1(\ldots), \ldots, \varphi_k(\ldots)), \quad w(t) = (w_1(t), \ldots, w_k(t))^*,
\]

where \( w_i(t) \) are independent standard Wiener processes.

It is well known that the generator of \( Z_s(t) \) is given by

\[
L_s(x, y) = \varepsilon^{-2} a_{00}(x, y) \frac{\partial^2}{\partial x^2} + 2 \varepsilon^{-1} \sum_{i=1}^{d} a_{i0}(x, y) \frac{\partial^2}{\partial x \partial y_i} + \sum_{i,j=1}^{d} a_{ij}(x, y) \frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^{d} b_i(x, y) \frac{\partial}{\partial y_i}
\]

\[
:= \varepsilon^{-2} L_1(x, y) + \varepsilon^{-1} L_2(x, y) + L_3(x, y)
\]

(2.2)

with

\[
A = (a_{ij})_{i,j=1,\ldots,d} = \frac{1}{2} \sigma \sigma^*, \quad a_{00} = \frac{1}{2} \sum_{i=1}^{k} \varphi_i^2, \quad a_{i0} = \frac{1}{2} \sum_{i=1}^{k} \varphi_i \sigma_{ij}.
\]

We suppose that the following conditions concerning the coefficients of (2.1) are satisfied.

(A1) The coefficients \( \varphi, b, \sigma \) are Lipschitz continuous in \( (x, y) \) and, for each \( x \), their derivatives in \( y \) up to and including second-order derivatives are bounded continuous functions of \( y \).

(A2) There are positive constants \( c_1, c_2, c_3 \) such that

\[
c_1 \leq a_{00}(x, y) \leq c_2,
\]

\[
\sum_{i=1}^{d} (a_{ii}(x, y) + b_i^2(x, y)) \leq c_3(1 + |y|^2).
\]

Write for convenience \( p(x, y) = (a_{00}(x, y))^{-1} \).
The function \( p(x, y) \) has a limit \( p^\pm(y) \) in \( \text{Česaro sense} \), as \( x \to \pm \infty \):

\[
\lim_{x \to \pm \infty} x^{-1} \int_0^x p(t, y) \, dt = p^\pm(y)
\]

uniformly in \( y \in \mathbb{R}^d \) and, moreover, also uniformly in \( y \in \mathbb{R}^d \),

\[
\lim_{x \to \pm \infty} x^{-1} \int_0^x D_y p(t, y) \, dt = D_y p^\pm(y),
\]

\[
\lim_{x \to \pm \infty} x^{-1} \int_0^x D_y^2 p(t, y) \, dt = D_y^2 p^\pm(y).
\]

Here and below \( D_y u \) and \( D_y^2 u \) are the gradient vector and the matrix of second derivatives in \( y \) of \( u \). The same notation is applied to the \( x \) variable.

We suppose further that the coefficients \( a_{ij}(x, y), b_i(x, y), i = 1, \ldots, d, j = 0, \ldots, d \) and all their derivatives in \( y \) up to the second order also have averages in \( x \), as \( x \to \pm \infty \) with the weight \( p(x, y) \). To write these conditions in a compact form we use the following notation: for any function \( K(x, y) \) having the limit in \( x \to \infty \) \( \text{Česaro sense} \),

\[
K^+(y) := \lim_{x \to \infty} x^{-1} \int_0^x K(t, y) \, dt, \quad K^-(y) := \lim_{x \to -\infty} x^{-1} \int_0^x K(t, y) \, dt;
\]

\[
K^\pm(x, y) := K^+(y)1_{\{x > 0\}} + K^-(y)1_{\{x \leq 0\}}.
\]

For a function \( K(x, y) \) we write \( K \in \mathcal{K} \) if \( K^\pm \) exists and

\[
x^{-1} \int_0^x K(t, y) \, dt - K^\pm(x, y) = (1 + |y|^2)\varepsilon(x, y),
\]

where the function \( \varepsilon \) is bounded and satisfies

\[
\lim_{|x| \to \infty} \sup_{y \in \mathbb{R}^d} |\varepsilon(x, y)| = 0. \tag{2.3}
\]

(B2) For \( i = 1, \ldots, d, \, j = 0, \ldots, d \), we have

\[
pb_i, Dy(pb_i), D_y^2(pb_i), \, pa_{ij}, Dy(pa_{ij}), D_y^2(pa_{ij}) \in \mathcal{K}.
\]

**Remark 2.1.** It is easy to see that the conditions A ensure the null-recurrence of the process \( X_{x,y}(t) \), defined for each \( y \) as the Markov process with generator \( L_3(x, y) \) starting at \( x \). Hence, the limit in (1.1), as a rule, does not exist.

As is mentioned in Introduction, the limit behavior of the process \( \tilde{Z}_\varepsilon(t)=(\varepsilon X_{\varepsilon}(t), Y_{\varepsilon}(t)) \) \( = (\tilde{X}_\varepsilon(t), \tilde{Y}_\varepsilon(t)) \) will be studied here. It follows from (2.1) that

\[
d\tilde{X}_\varepsilon(t) = \phi(\tilde{X}_\varepsilon(t)/\varepsilon, \tilde{Y}_\varepsilon(t)) \, dw(t),
\]

\[
d\tilde{Y}_\varepsilon(t) = b(\tilde{X}_\varepsilon(t)/\varepsilon, \tilde{Y}_\varepsilon(t)) \, dt + \sigma(\tilde{X}_\varepsilon(t), \tilde{Y}_\varepsilon(t)) \, dw(t), \tag{2.4}
\]

\[
\tilde{X}_\varepsilon(0) = \varepsilon x, \quad \tilde{Y}_\varepsilon(0) = y \tag{2.5}
\]

and the generator of \( \tilde{Z}_\varepsilon(t) \) is

\[
\tilde{L}_\varepsilon(x, y) = L_1(x/\varepsilon, y) + L_2(x/\varepsilon, y) + L_3(x/\varepsilon, y). \tag{2.6}
\]
Below we also consider the process (2.4) for the initial conditions

\[ \tilde{X}_x(0) = x, \quad \tilde{Y}_x(0) = y. \]  

(2.7)

Introduce the notation

\[ \tilde{a}_{ij}(x, y) = \frac{(a_{ij} p^\pm(x, y))}{p^\pm(x, y)}, \quad \tilde{b}(x, y) = \frac{(b p^\pm(x, y))}{p^\pm(x, y)}. \]  

(2.8)

\[ \tilde{A}(x, y) = (\tilde{a}_{ij}(x, y))_{i, j = 0, 1, ..., d}, \quad \tilde{\sigma}(x, y) = (\tilde{A}(x, y))^{1/2}. \]

Observe, in passing, that \( \tilde{a}_{00}(x, y) = (p^\pm(x, y))^{-1} \), and consider the Markov diffusion process \( \tilde{Z}(t) = (\tilde{X}_0(t), \tilde{Y}_0(t)) \), described by the stochastic differential equation

\[ d\tilde{Z}(t) = \tilde{b}(\tilde{Z}(t)) \, dt + \tilde{\sigma}(\tilde{Z}(t)) \, dw(t); \quad \tilde{Z}(0) = (x, y), \]  

(2.9)

with \( \tilde{b}(z) = \tilde{b}(x, y) = (0, \tilde{b}(x, y))^* \).

The coefficients of (2.9) are smooth functions only in each halfspace \( \{x > 0\} \) and \( \{x < 0\} \) and can have jumps at the hyperplane \( x = 0 \). This circumstance makes not self-evident the uniqueness even of a weak solution of (2.9), which is essential for the approach below. We list some known sufficient conditions guaranteeing the weak uniqueness. In 1–3 below \( \tilde{A} \) is assumed to be uniformly nondegenerate.

1. In the case \( p^+(y) = p^-(y) \), \( (a_{ij} p^+(y) = (a_{ij} p^-)(y), i, j = 0, 1, ..., d \), weak uniqueness follows from Stroock and Varadhan (1979), Chapter 6.

2. For \( d = 1 \) (the component \( Y_x \) is also one-dimensional) the uniqueness of the weak solution of (2.9) follows from Krylov (1969), see also exercise 7.3.4 in Stroock and Varadhan (1979).

3. It follows from Bass and Pardoux (1987) that weak uniqueness holds if the coefficients \( \tilde{a}_{ij}, \tilde{b}_i \) are constant in each halfspace \( \{x > 0\} \) and \( \{x < 0\} \).

We believe that weak uniqueness can be proved for essentially more general situations. Therefore, instead of consideration these special cases only, we introduce the following condition.

C. If problem (2.9) has a solution, it is weakly unique.

The main result of this paper is the following theorem.

**Theorem 2.2.** Let conditions A–C be satisfied and let \( \tilde{Z}_\varepsilon(t) = (\tilde{X}_\varepsilon(t), \tilde{Y}_\varepsilon(t)) \) be a solution of (2.4) and (2.7). Then the process \( \tilde{Z}_\varepsilon(t) \) converges in distribution to the Markov process \( \tilde{Z}_0(t) = (\tilde{X}_0(t), \tilde{Y}_0(t)) \), as \( \varepsilon \downarrow 0 \). Moreover, the solution of (2.4) and (2.5) converges in distribution to the solution of (2.9) with \( \tilde{Z}(0) = (0, y) \).

**3. Proof of Theorem 2.2**

To prove Theorem 2.2 we shall use several simple lemmas. In these lemmas \( c, c_i, k \) stand for generic positive constants and \( \alpha(x, y) \) stands for generic scalar, vector or matrix functions which are bounded and satisfy (2.3).
Lemma 3.1. Let a function \((x, y) \mapsto f(x, y) : \mathbb{R}^1 \times \mathbb{R}^d \to \mathbb{R}^1\) be a Borel function satisfying
\[
|f(x, y)| \leq c(1 + |y|^k).
\] (3.1)
Assume
\[
F(x, y) := x^{-1} \int_0^x f(t, y) \, dt = (1 + |y|^k)\alpha(x, y),
\] (3.2)
and let \(u_\varepsilon\) be a unique solution of the problem
\[
D_x^2 u_\varepsilon = f(x/\varepsilon, y), \quad u_\varepsilon(0, y) = D_x u_\varepsilon(0, y) = 0.
\] (3.3)
Then
\[
D_x u_\varepsilon(x, y) = x(1 + |y|^k)\alpha(x/\varepsilon, y), \quad u_\varepsilon(x, y) = x^2(1 + |y|^k)\alpha(x/\varepsilon, y).
\] (3.4)
If, in addition, for each \(x\), the function \(F\) has two continuous derivatives in \(y\) satisfying
\[
D_y F(x, y) = (1 + |y|^k)\alpha(x, y), \quad D_y^2 F(x, y) = (1 + |y|^k)\alpha(x, y),
\] (3.5)
then similar bounds for the derivatives in \(y\) hold:
\[
D_y u_\varepsilon(x, y) = x^2(1 + |y|^k)\alpha(x/\varepsilon, y), \quad D_y^2 u_\varepsilon(x, y) = x^2(1 + |y|^k)\alpha(x/\varepsilon, y),
\] (3.6)

Proof. Integrating the equation in (3.3) leads to \(D_x u_\varepsilon(x, y) = x F(x/\varepsilon, y)\). So (3.2) yields the first equation in (3.4). By integrating it, we get
\[
u_\varepsilon(x, y) = \varepsilon^2 \int_0^{x/\varepsilon} t F(t, y) \, dt = (1 + |y|^k)x^2 \left( \varepsilon^2 \int_0^{x/\varepsilon} \alpha(t, y) \, dt \right)
\] (3.7)
and the second equation in (3.4) follows. Similarly, by differentiating in \(y\) Eq. (3.7) and using (3.5), we obtain (3.6). □

Lemma 3.2. Let conditions A and B be satisfied. Let \(V_\varepsilon(x, y)\) be a solution to any one of the following problems (3.8) for \(i = 1, \ldots, d\) or (3.9) for \(i, j = 0, 1, \ldots, d\):
\[
a_{00}(x/\varepsilon, y)D_x^2 u = b_i(x/\varepsilon, y) - \tilde{b}_i(x, y); \quad u(0, y) = D_x u(0, y) = 0,
\] (3.8)
\[
a_{00}(x/\varepsilon, y)D_x^2 u = a_{ij}(x/\varepsilon, y) - \tilde{a}_{ij}(x, y); \quad u(0, y) = D_x u(0, y) = 0.
\] (3.9)
Then
\[
V_\varepsilon(x, y) = x^2(1 + |y|^2)\alpha(x/\varepsilon, y), \quad D_y V_\varepsilon(x, y) = x^2(1 + |y|^2)\alpha(x/\varepsilon, y),
\] (3.10)
\[
D_y^2 V_\varepsilon(x, y) = x^2(1 + |y|^2)\alpha(x/\varepsilon, y), \quad D_x V_\varepsilon(x, y) = x(1 + |y|^2)\alpha(x/\varepsilon, y),
\] (3.10)
\[
D_x D_y V_\varepsilon(x, y) = x(1 + |y|^2)\alpha(x/\varepsilon, y).
\] (3.10)

Here, as before, \(\alpha(x, y)\) are various bounded functions with property (2.3).
Proof. Upon dividing Eqs. (3.8) and (3.9) by $a_{00}(x, \varepsilon, y)$ we rewrite them in the form (3.3). Note now that by definition
\[ \bar{a}_{ij}(x, y) = \bar{a}_{ij}(x, \varepsilon, y), \quad p^\pm(x, y) = p^\pm(x, \varepsilon, y), \] etc.
To prove (3.10) for the solution of (3.9) it is enough to check the conditions (3.1), (3.2), and (3.5) for the function \( \varphi(x, y) = p(x, y)(a_{ij}(x, y) - \bar{a}_{ij}(x, y)) \).
Condition (3.1) follows from A2. Further, taking into account conditions B1 and B2, we have that, for \( x \geq 0 \),
\[ N_{BS}(x, NSI; y) = x - \frac{1}{x} \int_0^x \varphi(t, \varepsilon, y) dt \]
\[ = x - \frac{1}{x} \int_0^x p(t, \varepsilon, y) a_{ij}(t, \varepsilon, y) dt \]
\[ - (a_{ij} p)^+(y) - (a_{ij} p)^+(y) \left[ 1 - (p^+(y))^{-1} x^{-1} \int_0^x p(t, \varepsilon, y) dt \right] \]
\[ = (1 + |y|^2) \chi_1(x, \varepsilon, y) + (1 + |y|^2) \chi_2(x, \varepsilon, y) \]
and hence (3.2) holds as well. In the case \( x < 0 \), the proof is exactly the same. Let us check now conditions (3.5). The condition \( |\varphi'(x, y)| < c \) follows from A1. Then, confining ourselves for brevity only to the case \( x > 0 \) leads to
\[ D_y \Phi(x, y) = x^{-1} \int_0^x D_y \varphi(t, y) dt \]
\[ = x^{-1} \int_0^x D_y[ p(t, y)(a_{ij}(t, y) - \bar{a}_{ij}(t, y))] dt \]
\[ = x^{-1} \int_0^x D_y[ p(t, y)(a_{ij}(t, y)) dt - D_y((a_{ij} p)^+(y)) \]
\[ + D_y \left[ (a_{ij} p)^+(y)(p^+(y))^{-1} (p^+(y) - x^{-1} \int_0^x p(t, y) dt) \right] \]
\[ = (1 + |y|^2) \chi(x, y) \]
due to conditions B1, B2. Verification of (3.5) for $D_y^2 \varphi$ and $D_y^2 \Phi$ is completely similar. In exactly the same way one can prove (3.10) for solutions of (3.8). \( \square \)

Lemma 3.3. Take a \( \varepsilon > 0 \) and let \((\tilde{X}_\varepsilon(t), \tilde{Y}_\varepsilon(t))\) be a solution of the problem (2.4) and (2.7). Then, for any \( k \geq 1 \), there is a constant \( c_k \) independent of \( \varepsilon \) and such that, for \( t \geq 0 \),
\[ E \left\{ \sup_{s \leq t} \left| \tilde{X}_\varepsilon(s) \right|^2 + \left| \tilde{Y}_\varepsilon(s) \right|^2 \right\} \leq (|x|^2k + |y|^{2k} + 1) \exp\{c_k t\}. \]
(3.11)
For any \( T > 0 \), the inequality
\[ E|\tilde{X}_\varepsilon(t + h) - \tilde{X}_\varepsilon(t)|^4 + E|\tilde{Y}_\varepsilon(t + h) - \tilde{Y}_\varepsilon(t)|^4 \leq c h^2 \]
(3.12)
holds if \( t, t + h \in [0, T] \) with constant \( c \) independent of \( t, h, \) and \( \varepsilon \).
Proof. It is easy to see from conditions A2 that
\[
\left( \frac{\partial}{\partial t} + \tilde{L}_e \right) \left[ (|x|^{2k} + |y|^{2k} + 1) \exp\{-c_k t\} \right] \leq 0
\]
for \(c_k\) sufficiently large. It guarantees that the process
\[
(|\tilde{X}_e(t)|^{2k} + |\tilde{Y}_e(t)|^{2k} + 1) \exp\{-c_k t\}
\]
is a local supermartingale, and (3.11) follows. The inequality (3.12) follows from (3.11) and properties of stochastic integrals.

Remark 3.4. Lemma 3.3 guarantees compactness of the distributions of the processes \((\tilde{X}_e(t), \tilde{Y}_e(t))\) in \(C[0,T]\) for any \(T < \infty\), see Prokhorov, 1956.

Let \(\mathcal{F}_t\) be the monotone family of \(\sigma\)-algebras generated by \(\tilde{Z}_e(s)\) for \(0 \leq s \leq t\).

Lemma 3.5. For any \(T < \infty\), \(i = 1, \ldots, d\), and \(j, k = 0, 1, \ldots, d\), we have
\[
\sup_{t \leq T} \left| \int_0^t (b_i(\tilde{X}_e(s)/\varepsilon, \tilde{Y}_e(s)) - \tilde{b}_i(\tilde{X}_e(s), \tilde{Y}_e(s))) \, ds \right| \to 0,
\]
\[(3.13)\]
\[
\sup_{t \leq T} \left| \int_0^t (a_{jk}(\tilde{X}_e(s)/\varepsilon, \tilde{Y}_e(s)) - \tilde{a}_{jk}(\tilde{X}_e(s), \tilde{Y}_e(s))) \, ds \right| \to 0
\]
\[(3.14)\]
as \(\varepsilon \downarrow 0\) in probability.

Proof. The proofs of assertions (3.13) and (3.14) are similar, so we prove only the first one. Let \(V_\varepsilon\) be the solution of (3.8). It has two derivatives in \((x, y)\), which are perhaps discontinuous only at \(x = 0\). Due to results of Krylov (1977) and nondegeneracy condition in A2 one can still apply Itô’s formula to \(V_\varepsilon(\tilde{X}_e(t), \tilde{Y}_e(t))\) and get
\[
V_\varepsilon(\tilde{X}_e(t), \tilde{Y}_e(t)) = V_\varepsilon(x, y)
\]
\[
+ \int_0^t (b_i(\tilde{X}_e(s)/\varepsilon, \tilde{Y}_e(s)) - \tilde{b}_i(\tilde{X}_e(s), \tilde{Y}_e(s))) \, ds + A_t + B_t + M_t,
\]
\[(3.15)\]
where, for \(y_0 := x\),
\[
A_t = \sum_{j \neq k > 0} \int_0^t a_{ij}(\tilde{X}_e(s)/\varepsilon, \tilde{Y}_e(s)) \left( \frac{\partial^2 V_\varepsilon}{\partial y_i \partial y_j}(\tilde{X}_e(s), \tilde{Y}_e(s)) \right) \, ds,
\]
\[
B_t = \sum_{j=1}^d \int_0^t b_j(\tilde{X}_e(s)/\varepsilon, \tilde{Y}_e(s)) \left( \frac{\partial V_\varepsilon}{\partial y_j}(\tilde{X}_e(s), \tilde{Y}_e(s)) \right) \, ds,
\]
\[
M_t = \int_0^t \left( \frac{\partial V_\varepsilon}{\partial \tilde{X}}(\tilde{X}_e(s), \tilde{Y}_e(s)) \phi(\tilde{X}_e(s)/\varepsilon, \tilde{Y}_e(s)) \right) \, dw_s
\]
\[
+ \int_0^t D_y V_\varepsilon(\tilde{X}_e(s), \tilde{Y}_e(s)) \sigma(\tilde{X}_e(s), \tilde{Y}_e(s)) \, dw_s.
\]
It follows from (3.10) that \( V_\varepsilon(x, y) \to 0 \). By using again (3.10) and also Lemma 3.3 and bearing in mind the equality

\[
1 = I_{\{|X_\varepsilon(t)| \geq \sqrt{\varepsilon}\}} + I_{\{|X_\varepsilon(t)| < \sqrt{\varepsilon}\}},
\]

we obtain

\[
E \sup_{t \leq T} |V_\varepsilon(X_\varepsilon(t), Y_\varepsilon(t))| = E \sup_{t \leq T} |X_\varepsilon^2(t)| (1 + |Y_\varepsilon(t)|)^k \varphi(X_\varepsilon(t)/\varepsilon, Y_\varepsilon(t)) |x| \leq \sqrt{\varepsilon}
\]

\[
\leq N \sup_{|x| \geq \sqrt{\varepsilon}} |x(x, y)| + M \varepsilon \to 0.
\]

In the same way,

\[
E \sup_{t \leq T} |A_t| \leq N \sup_{t \leq T} (1 + |Y_\varepsilon(t)|^2) |D_z D_y V_\varepsilon(X_\varepsilon(t), Y_\varepsilon(t))| + |D_y^2 V_\varepsilon(X_\varepsilon(t), Y_\varepsilon(t))| \to 0, \quad E \sup_{t \leq T} |B_t| \to 0.
\]

It follows from (3.15) that to prove (3.13) it only remains to prove that \( \sup_{t \leq T} |M_t| \to 0 \) in probability. This is known to be equivalent to proving the convergence of the quadratic variations of the martingales \( M_t \) to zero. The latter is done following the same pattern as above.

Now we can prove the theorem. First we outline the main ideas. Notice that the process \((\tilde{X}_\varepsilon(t),\tilde{Y}_\varepsilon(t)) - \int_0^t b(\tilde{X}_\varepsilon(s)/\varepsilon, \tilde{Y}_\varepsilon(s)) \, ds\) is a martingale. Due to Lemma 3.3 by utilizing Skorokhod’s theorem from [13], for any sequence \( \varepsilon_n \) going to zero one can choose a subsequence \( \varepsilon_n' \) such that the following holds:

1. There are processes \( \tilde{Z}_{\varepsilon_n'}(t) \), defined on some probability space having the same finite-dimensional distributions as \( \tilde{Z}_{\varepsilon_n}(t) \).

2. \( \tilde{Z}_{\varepsilon_n'}(t) \to \tilde{Z}(t) = (\tilde{X}(t), \tilde{Y}(t)) \) in probability, where \( \tilde{Z}(t) \) is an a.s. continuous stochastic process.

Further, similarly to the proof of Lemma 4.1 in Khasminskii (1968) one can establish that \((\tilde{X}(t), \tilde{Y}(t) - \int_0^t b(\tilde{X}(s), \tilde{Y}(s)) \, ds)\) is a martingale, and \( \tilde{Z}(t) \) satisfies Eq. (2.9).

An implementation of this argument in a much more general setting can be found in Liptser and Shiryaev, 1989 (see Theorem 8.3.3 there) under the additional assumption that \( \tilde{a} \) and \( \tilde{b} \) are continuous. In our situation, generally speaking these coefficients are discontinuous at \( x = 0 \), which is a set of Lebesgue measure zero. However, the \( x \) component of our processes is uniformly nondegenerate and this allows one to carry out the program in our case as well. The details of passing to the limit in the case of coefficients with Lebesgue measure zero of the set of their discontinuity can be found in Chao, Yi-Ju (1999). Now, the assertion of Theorem follows from Condition C.

4. Corollary and example

Theorem 2.1 provides some information on the limit behavior of solutions to elliptic and parabolic differential equations with different scales of variables. We formulate here one result of this type. Below by \( W^{2,1}_{d+1,\text{loc}} \) we mean the space of functions \( u(x, y, t) \) on
[0, ∞) × \mathbb{R}^{d+1} which are bounded and continuous and have generalized derivatives of first and second orders in (x, y) variables, as well as the first generalized derivative in t, locally summable to the power d + 1 in (0, ∞) × \mathbb{R}^{d+1}.

**Corollary 4.1.** Let the coefficients of the operator \( L_n(x, y) \) satisfy conditions A and B and let the matrix \((a_{ij})_{i,j=0}^d\) be uniformly nondegenerate. Denote

\[
L(x, y) = \sum_{i,j=0}^d a_{ij}(x, y) \frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^d b_i(x, y) \frac{\partial}{\partial y_i},
\]

\[
\tilde{L}(x, y) = \sum_{i,j=0}^d \tilde{a}_{ij}(x, y) \frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^d \tilde{b}_i(x, y) \frac{\partial}{\partial y_i}
\]

(here, as before, we denote \( x = y_0 \), the coefficients \( \tilde{a}_{i,j}, \tilde{b}_i \) are defined in (2.8)). Also assume that, for any bounded and infinitely differentiable function \( N(x, y) \), the problem

\[
\frac{\partial u}{\partial t} = \tilde{L}(x, y)u; \quad u(x, y, 0) = \phi(x, y)
\]

has a unique bounded solution \( \tilde{u}(x, y, t) \in W^{2,1}_{d+1, \text{loc}} \). Then, for any bounded and infinitely differentiable function \( \phi(x, y) \), bounded solutions \( u_\varepsilon(x, y, t) \in W^{2,1}_{d+1, \text{loc}} \) of the problem

\[
\frac{\partial u_\varepsilon}{\partial t} = L(x/\varepsilon, y)u_\varepsilon; \quad u_\varepsilon(x, y, 0) = \phi(x, y)
\]

satisfy

\[
\lim_{\varepsilon \to 0} u_\varepsilon(x, y, t) = \tilde{u}(x, y, t).
\]

The proof follows immediately from the well-known probabilistic representations of \( u_\varepsilon(x, y, t) \) and \( \tilde{u}(x, y, t) \) (see, e.g. Section 2.10 of Krylov, 1977), Theorem 2.2, and the fact that condition C is equivalent to uniqueness of solution to (4.1).

**Remark 4.1.** One can get similar results for inhomogeneous parabolic equations. For instance, let \( f(x, y) \) be a function such that the function \( f(x, y) p(x, y) \) and its derivatives in \( y \) up to and including the second order derivatives have limits in Cesaro sense for \( x \to \pm \infty \) uniformly in \( y \). Denote

\[
\tilde{f}(x, y) = \frac{(fp)^+(y)}{p^+(y)} \mathbf{1}_{\{x > 0\}} + \frac{(fp)^-(y)}{p^-(y)} \mathbf{1}_{\{x \leq 0\}}.
\]

Also assume that the conditions of Corollary hold. Then the solutions \( v_\varepsilon(x, y, t) \in W^{2,1}_{d+1, \text{loc}} \) of the problem

\[
\frac{\partial v_\varepsilon}{\partial t} = L(x/\varepsilon, y)v_\varepsilon + f(x/\varepsilon, y); \quad v_\varepsilon(x, y, 0) = \phi(x, y)
\]

converge for \( \varepsilon \to 0 \) to the solution \( \tilde{v}(x, y, t) \in W^{2,1}_{d+1, \text{loc}} \) of the problem

\[
\frac{\partial \tilde{v}}{\partial t} = \tilde{L}(x, y)\tilde{v} + \tilde{f}(x, y); \quad \tilde{v}(x, y, 0) = \phi(x, y).
\]

The proof follows from the probabilistic representation of \( v_\varepsilon(x, y, t) \) and the arguments used in Section 3.
Example 4.1. Consider the simplest situation where the conditions of Theorem 2.1 are satisfied, \( p^\pm, (bp)^\pm, (a_{11} p)^\pm, (a_{01} p)^\pm \) are constants, and
\[
(a_{01} p)^\pm = 0; \quad (a_{11} p)^+ = (a_{11} p)^- = a_{11} p.
\]
\[\text{(4.2)}\]

In this case, Eq. (2.9) can be solved explicitly. So we have from Theorem 2.1: the solution \((X_t, Y_t)\) of (2.1) has the following limit behavior:
\[
\begin{align*}
\varepsilon X_t & \xrightarrow{\text{distr}} X_0(p^+, p^-, t) := X_0(t), \quad Y_t \xrightarrow{\text{distr}} \left(\frac{bp^+}{p^+} - t + \frac{bp^+}{p^+} - \frac{bp^-}{p^-}\right) \int_0^t 1_{\{X_s > 0\}} \, ds + \left(\frac{2a_{11} p}{p}\right)^{1/2} w_1(t).
\end{align*}
\]

Here \(X_0(t)\) is a unique continuous solution of the SDE
\[
dX(t) = [(2/p^+)^{1/2} 1_{\{X(s) > 0\}}] + (2/p^-)^{1/2} 1_{\{X(s) \leq 0\}} \, dw(t), \quad X(0) = 0,
\]
\(w(t), w_1(t)\) are independent standard Brownian motions. It is easy to see from Portenko (1990, p. 143) that the transition probability density for \(X_0(t)\) can be written as follows:
\[
G_{a^+, a^-}^+(t, x, y) = \begin{cases}
(2a_-)^{-1/2} G_c(t, x/\sqrt{2a_-}, y/\sqrt{2a_-}) & \text{if } x < 0, \quad y < 0, \\
(2a_+)^{-1/2} G_c(t, x/\sqrt{2a_+}, y/\sqrt{2a_+}) & \text{if } x < 0, \quad y > 0, \\
(2a_-)^{-1/2} G_c(t, x/\sqrt{2a_-}, y/\sqrt{2a_-}) & \text{if } x > 0, \quad y < 0, \\
(2a_+)^{-1/2} G_c(t, x/\sqrt{2a_+}, y/\sqrt{2a_+}) & \text{if } x > 0, \quad y > 0.
\end{cases}
\]

Here
\[
G_c(t, x, y) = g(t, x - y) + 1_{\{y > 0\}} cg(t, |x| + |y|), \quad c = \frac{\sqrt{a_+} - \sqrt{a_-}}{\sqrt{a_+} + \sqrt{a_-}}
\]
and \(g(t, x - y)\) is the transition probability density of standard Brownian motion.

It is also known (see Watanabe, 1999; Khasminskii, 1999) that the random variable \(\eta(t) = \int_0^t 1_{\{X(s) > 0\}} \, ds\) has the generalized Arcsine distribution: for \(x \geq 0\)
\[
P\{\eta(t)/t < x\} = 2/\pi \sin^{-1} \sqrt{\frac{p^- x}{p^+ + (p^- - p^+) x}}.
\]

So we see, in particular, that, up to some constants, the distribution of the slow component for fixed \(t\) in this example converges to the convolution of the generalized Arcsine and Gaussian distributions if \((bp)^+/p^+ \neq (bp)^-/-p^-\).

References


